Life Cycle Management (Subject Editor: Gerald Rebitzer)

Fast Life Cycle Assessment of Synthetic Chemistry (FLASC™) Tool

Alan D. Curzons[1], Concepción Jiménez-González[2]*, Ailsa L. Duncan[3], David J. C. Constable[4] and Virginia L. Cunningham[4]

[1]GSK, CEHS, Southdownview Way, Worthing BN14 8NQ, UK
[2]GSK, CEHS, CS.1152, Five Moore Dr, Research Triangle Park, NC, 27709-3398, USA
[3]GSK, CEHS, GSK House, 980 Great Western Road, Brentford, TW8 9GS, UK
[4]GSK, CEHS, 1 Franklin Plaza, Philadelphia, PA 19103, USA

*Corresponding author (conchita.j.gonzalez@gsk.com)

DOI: http://dx.doi.org/10.1065/lca2007.03.315

Please cite this paper as: Curzons AD, Jiménez-González C*, Duncan AL, Constable DJC, Cunningham VL (2007): Fast Life Cycle Assessment of Synthetic Chemistry (FLASC™) Tool. Int J LCA 12 (4) 272–280

Abstract

Background, Aim and Scope. There is a clear need for simple methodology to deliver metrics that may be used to determine and benchmark the 'greenness' or relative sustainability of synthetic processes for Active Pharmaceutical Ingredients (APIs). Such methodology and metrics should facilitate more informed and sustainable business choices. This capability is particularly important at an early stage in R&D development activities when route and processes are being selected and detailed environmental data are not available. FLASC™ (Fast Life cycle Assessment of Synthetic Chemistry) is a web-based tool and methodology designed to meet these requirements.

Materials and Methods. FLASC™ was developed from a detailed assessment of the cradle-to-gate life cycle environmental impacts associated with the manufacture of materials used in a typical pharmaceutical process.

Results. This paper describes the methodology used to develop FLASC™ and provides examples of the type of information and guidance FLASC™ provides.

Discussion. Both Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA) were used for the statistical analysis during the development of FLASC™. Benchmarking within the pharmaceutical industry and use of normalization for molecular complexity were also integrated to the tool.

Conclusions. FLASC™ represents an important part of the overall efforts of GlaxoSmithKline (GSK) to incorporate and maintain sustainable business practices for manufacture of APIs used in its pharmaceutical products.

Recommendations and Perspectives. This tool is not intended to assess waste from GSK operations nor solvent recovery and currently does not incorporate specific chemical-related health and safety data. However, these are already routinely assessed within GSK R&D at appropriate milestones and the use of FLASC™ is complementary to these evaluations.

Keywords: Green chemistry; LCA of pharmaceuticals; principal component analysis; synthetic chemistry

1 Background, Aim and Scope

How do you identify the greenest process to an API?

GSK has already developed an Eco-Design Toolkit©. It is a web-based suite of tools and methodologies that provides concise, practical, and simple information and guidance to scientists and engineers. The Eco-Design Toolkit© is intended to facilitate selection of better materials, greener chemistries, and design of greener processes through a strong focus on effective resource (mass and energy) utilisation.

While these existing tools have proven to be very valuable, a tool that would allow scientists and engineers to easily compare the 'greenness' of their processes has been missing. With increasing technical and business demands on scientists, shorter timelines, and ever more stringent regulatory scrutiny, there is a strong need to make comparisons and decisions earlier in the R&D process, before Environmental, Health or Safety data are generally available. Consequently, there is a significant benefit from any approach that provides an early understanding and measure of process 'greenness' with the ability to benchmark performance.

2 A Life Cycle Based Approach

It has been the premise of GSK that the use of life cycle inventory and assessment techniques could deliver a simple yet robust method for assessing and comparing process 'greenness'. Previous GSK studies have demonstrated that there are significant benefits from using life cycle based approaches that organise impact information around a set of commonly accepted 'sustainability' metrics. A method for achieving this was developed and is reported elsewhere. [1–3].

A successful and practical tool should have the following elements:

- ability to measure the environmental life cycle impacts and GSK operational impacts; i.e., a true measure of the 'greenness' of GSK processes;
- facility for understanding the environmental impacts GSK processes cause combined with insights into the causes of high impacts and guidance for how to reduce these impacts;
- simplicity and ease of use;
- relevant, meaningful, accurate and easily understood information that is readily available to scientists.

FLASC™ was developed to integrate these principles.

3 Methodology

The following paragraphs describe the methodology for developing FLASC™ within GSK.

3.1 Overview

In order to undertake a life cycle environmental impact assessment of any chemical synthetic route or process, it is essential to obtain life cycle data for all the materials in that process. However, it has been a difficult and frustrating task to obtain publicly or commercially available LCI/A data for materials of interest to the pharmaceutical industry. Consequently, a GSK LCI/A program was undertaken to develop a fast, streamlined approach that delivers credible life cycle assessments for a wide range of materials commonly used in drug manufacture.

A modular approach described elsewhere [4,5] was used to generate life cycle inventory (LCI) data for approximately 140 materials. These raw LCI data were collated using the eight core GSK 'sustainability metrics' impact categories [3]. Hierarchical cluster and principal components analysis were used to group materials that had similar impact profiles, and from these groups, a simple classification process was developed. A total of 14 unique material classes were identified from this data set.

For each of these 14 material classes it was possible to generate average life cycle impact profile data that could be used for materials where LCI data did not exist. A methodology was then developed to predict the cradle-to-gate life cycle impact profile for the typical batch chemical process used to synthesise a GSK API. This methodology was based on the LCI of the materials used in the process, using a combination of actual or average data, and the mass of the material.

This approach was used to generate a core set of life cycle impact profiles for 22 well-developed GSK processes to APIs. These 22 processes represent approximately 84 batch chemical operations that may contain one or more chemical transformations, followed by separation and/or isolation steps. This core data set was then used to develop a series of formulae that enabled a score to be calculated for each of the eight impact categories. An average score, termed the FLASC™ score, was then calculated from the individual scores for each impact category.

A GSK intranet site has now been developed to exploit this novel approach to generate cradle-to-gate life cycle assessments for batch chemical processes typically found in the pharmaceutical industry. The intranet site enables users in GSK R&D and manufacturing operations to evaluate and compare new synthetic routes and benchmark against existing GSK processes. In addition, materials in the route (and/or process) that have the greatest cradle-to-gate life cycle environmental impacts are identified, data for mass productivity and reaction mass efficiency are provided, and general guidance is offered. The guidance is provided so that future route development activities can focus on areas that will have the greatest influence on improving the 'greenness' of the process.

Detailed analysis and validation of the tool indicate that as long as the rules described within this paper are applied, the errors associated with this approach are relatively small.

3.2 Generating life cycle data for process materials

The methodology and heuristics to generate life cycle inventory/assessment (LCI/A) data for materials have been reported elsewhere [4,5]. This approach generates discrete gate-to-gate life cycle inventories using standard chemical engineering process design principles. Each discrete gate-to-gate module may be linked in any of a variety of production chains, to provide a full cradle-to-gate life cycle inventory.

The boundaries for each cradle-to-gate LCI included the extraction, production and transport of raw materials; energy production for the entire cradle-to-gate; and the manufacture of the final chemical. For transportation distances and modes, average US reported data were used [6]. The raw LCI information for each chemical is organised into the following categories: raw materials, energy requirements, air emissions, water emissions and solid waste.

In addition, the inventory is broken down into contributions to the life cycle from energy use, manufacturing and transportation.

Approximately 140 cradle-to-gate life cycle inventories of chemicals were generated as described above and constitute the base data set for the development of this tool.

3.3 Life cycle assessment applying GSK metrics

The approach described above enables the life cycle environmental impact assessment values for materials to be organised around a set of commonly accepted 'sustainability' metrics. Life cycle impact assessment values were determined for the following eight impact categories:

- Net Mass of materials used [kg];
- Energy required [MJ];
- Green House Gas Equivalents [GHG, kg of CO_2-equivalents];
- Oil and natural gas depletion for materials manufacture [kg];
- Acidification Potential [AP, kg of SO_2 equivalents];
- Eutrophication Potential [EP, kg of $(PO_4)^{-3}$ equivalents];
- Photochemical Ozone Creation Potential [POCP, kg of ethene-equivalents];
- Total Organic Carbon (TOC) load before waste treatment.

Oil and natural gas depletion does not include oil and natural gas used for energy generation, but only the resources used as feedstock for material manufacture. The total organic carbon data represents the pre-treatment carbon loading, which is subsequently evaluated using common wastewater treatment models once FLASC results are obtained.

Boundaries. The following boundaries have been applied:

- Emissions, energy and material consumption resulting from energy production, and transportation are included in the final material cradle-to-GSK-gate LCIs and LCAs;
- Since the tool is designed to benchmark the relative 'greenness' of processes used to synthesise APIs, life cycle environmental impacts from packaging materials are excluded;
- Life cycle environmental impacts associated with distribution and final fate are not within the scope of this program.

3.4 Statistical analysis and materials grouping (Pirouette[R] analysis)

Once LCI cradle to gate data were developed for all the materials around the 8 impact categories, it was possible to employ related statistical techniques such as Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) to evaluate relationships that existed between materials having similar environmental impact profiles. In other words, PCA and HCA objectively verify statistically consistent groupings across multiple environmental impact categories. This analysis of the impact data was undertaken in partnership with Infometrix using their proprietary software, Pirouette[R] [7].

Pirouette[R] allows users to very rapidly analyze large data sets using various multivariate statistical techniques and visualise how data are clustered and related. It also permits the development of models that can be used to predict or construct data sets based on the training set of life cycle impacts. Both Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA) were used for data analysis. A detailed discussion of these multivariate statistical methods may be found elsewhere [8–10]. Principal Components Analysis showed that the impact category data from all eight categories could be sufficiently described by or reproduced from three principal components. These principal components are essentially composite vectors (composed of multiple impact categories) in the data space. Data points plotted (3 dimensionally) against these principal components will generally form discrete clusters based on similarities in life cycle environmental impacts. For example, Fig. 1 shows a plot of substances that cluster, based on the life cycle impact differences between lithium and non-lithium based inorganic metal salts. Fig. 2 shows a principal components scores plot of the environmental life cycle impacts. This plot shows the result one might expect in that for the entire data set there are greater similarities between GHG equivalents, gross energy and oil equivalents than there are between mass, POCP, TOC and eutrophication. HCA plots show clustering of related materials into discrete groups.

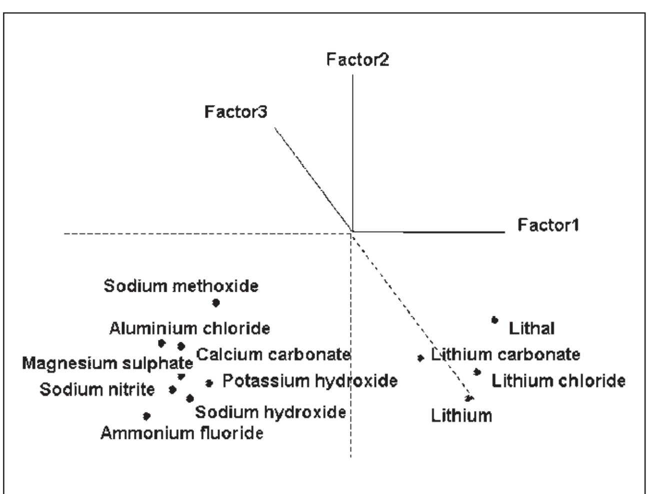

Fig. 1: Shows an example of PCA for selected inorganic materials. The figure illustrates substances that cluster based on the life cycle impact differences between lithium and non-lithium based inorganic metal salts, therefore illustrating the suitability of the classifications to estimate life cycle impacts of unknown substances. This representation is achieved by putting all the variables on equal statistical footing: zero means and values expressed in terms of variance

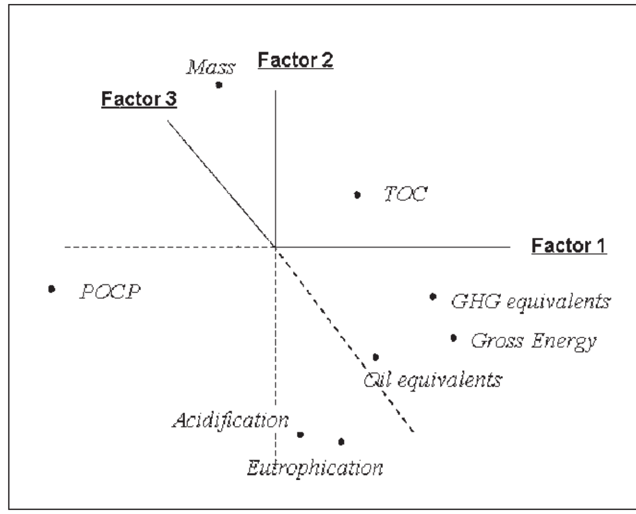

Fig. 2: An example of PCA for metrics. The figure shows a principal components scores plot of the environmental life cycle impacts. This plot shows a result one might expect in that for the entire data set there are greater similarities between GHG equivalents, gross energy and oil equivalents than there are between mass, POCP, TOC and eutrophication

Following the discovery that HCA and PCA could in fact be used to easily group materials with related impact profiles, a number of approaches were evaluated to organise grouped materials according to logical classifications. For example, materials could be classified by functional group (e.g., an alcohol, ether or ketone) but this was found to be unworkable for complex molecules (i.e., multiple functional groups or heteroatoms) and did not sufficiently discriminate among many of the materials. A simple approach was finally developed, where all 140 materials in the data set could be grouped into 14 relatively straightforward classes. An underlying premise of this approach is that within a particular class, the life cycle environmental impacts for the materials in that class are similar. The 14 classes may be further organised into three major groups as shown in **Table 1**.

Once materials were grouped into logical classes, the next step was to ensure that for any new materials used by GSK, consistent and meaningful classifications could be assigned. While many classifications were straightforward, where ambiguity is encountered, a simple set of heuristics based on molecular weight cut-off was adopted to enable the correct classification. A detailed evaluation of materials used in 35 of the most developed processes passing through GSK R&D during the period spanning from 1990 to 2000 revealed that, using this approach, all the materials in those processes (250 materials in total) could be readily and unambiguously classified.

Table 1: Categorisation of materials . The table below presents the 14 categories used in FLASC and its definitions

Organic	**Aliphatic.** Simple aliphatic compounds with molecular weight of less than 110, excluding halogens.
	Alkane/alkenes/aklynes. Containing C and H only. This classification includes branched compounds. It is assumed that long chain hydrocarbons are rarely used in late stage pharmaceutical manufacturing so no molecular weight limit is included in the Definition.
	Mono-substituted aromatic. Contains an aromatic ring, with no more than one substitute carbon and the molecular weight less than 140, excluding halogens.
	Poly-substituted aromatic. Contains an aromatic ring, has greater than one substituted carbon and the molecular weight is less than 140, excluding halogens.
	Complex organic. Will include :
	• aliphatics with a molecular weight greater than 110, excluding halogens;
	• aromatics with a molecular weight of greater than 140, excluding halogens;
	• heterocycles with a molecular weight of greater than 110, excluding halogens;
	• must not have a molecular weight greater than 220 excluding halogens.
	Pyridine derivative. Materials with pyridine
	Simple heterocycle. Heterocycle with a molecular weight of less than 110, excluding halogens.
	Speciality material. materials with a molecular weight greater than 220, excluding halogens. Natural products.
	Complex intermediate. Pharmaceutical intermediates in synthetic routes with molecular weight greater than 500.
Inorganic	**Contains lithium.** Any material containing lithium.
	Contains sulphur. Any inorganic material containing sulphur unless it contains lithium.
	Contains a metal cation. Any material containing a metal cation other than lithium. Ammonium is regarded as a metal cation.
	General. Any inorganic that does not fit in the categories above.
Solvent	**Single class for all solvents.** Data for most solvents used in the company are included in the tool.

3.5 Benchmarking and development of scoring process

The same 35 processes discussed in section 3.4 were evaluated to remove those that were considered to be outside the norm for standard batch chemical operations used to manufacture APIs in the pharmaceutical industry. This left a core training set of 22 GSK pharmaceutical processes that have been run at scale, either in a GSK Pilot Plant or in a production facility. These processes are not intended to be representative of all batch chemical processes that exist in the chemical industry, but they are representative of the GSK synthetic chemistry process practices and by extension, representative of current pharmaceutical industry practices. The number of process stages in a given process varied from 3 to 12 with an average of 7 stages being common for the pharmaceutical industry. Each stage generally results in an isolated intermediate, although there may be several chemical transformations or process steps in a given stage.

Process description reports were used to identify all materials used in each process and to determine the mass used (expressed as kg used per kg API produced). Cradle-to-gate life cycle data were obtained for each material in one of two ways:

- for a material already in the database, actual life cycle impact data were used;
- for new materials not in the database, once classified, average life cycle impact data for the class were used.

The overall cradle-to-gate life cycle impact was then calculated for processes/routes by multiplying the mass of each material by the life cycle impact value for that material and summing the data for a given impact category across all the materials used in the process. This is represented as shown in Eq. 1:

$$C_i = \sum_{j=1}^{N} m_j c_{ij} \qquad (1)$$

where:

C_I = Value of life cycle impact category i for the route under study.
i = Life Cycle impact category (e.g. net mass, gross energy, GHG, etc.)
j = material (e.g. acetone, ethanol, etc.)
c_{ij} = Value of life cycle impact category i for the material j
m_j = Mass of the material j
N = number of materials used in a process/route

This approach was used for each of the 22 processes/routes and provided the benchmark data set. Each of the 22 processes/routes contained a summed life cycle value for each of the eight impact categories. This benchmark data set was then used to develop a simple scoring approach that enables new processes/routes to be assessed and compared.

A normalization step was employed using a logarithmic approach. The logarithmic approach was chosen to normalise each impact category into a 1–5 scale in view of the large range of values of the impact categories in the training set. The normalisation approach is shown in **Eq. 2**. The equation is bounded by the upper and lower limits of the data for each impact category, where these limits may be thought of as the range of life cycle 'performance' for each impact category.

Using these formulae, it was possible to derive a simple score for each impact category for a given process/route based on its life cycle impact data. The principle is that the log 10 of the maximum or worst value within an impact category scores 1 and that the log 10 of the minimum or best value within an impact category scores 5. Currently, if the value is below the lower environmental impact limit, the maximum score remains 5. Likewise, if the value is above the upper environmental impact limit, the minimum score remains 1. The final score (FLASC™ score) is the mean of the scores derived for each of the 8 impact categories. There-

fore, the greener the process the higher the associated FLASC™ score (**Eq. 3**).

$$s_i = 4\left(\frac{Log(M_i) - Log\left[C_i \dfrac{\overline{MW}}{mw}\right]}{Log(M_i) - Log(m_i)}\right) + 1 \tag{2}$$

$$If \begin{cases} \overline{s_i} < 1 \Rightarrow S = 1 \\ 1 < \overline{s_i} < 5 \Rightarrow S = \overline{s_i} = \dfrac{\sum_{i=1}^{8} s_i}{8} \\ \overline{s_i} > 5 \Rightarrow S = 5 \end{cases} \tag{3}$$

where:

i = Life Cycle impact category (e.g. net mass, gross energy, GHG, etc.);
m_i = minimum value of life cycle impact category i for the benchmark data set;
M_i = maximum value of life cycle impact category i for the benchmark data set;
s_i = value of life cycle category i for material j;
$\overline{s_i}$ = arithmetic mean of the scores for the eight metrics;
S = FLASC™ score for the route evaluated;
mw = molecular weight of the final product for the route evaluated;
$\overline{MW}$ = average molecular weight for the benchmark data set.

There is considerable debate in the literature regarding the weighting of environmental impacts and the reader is referred elsewhere for a discussion of this issue [11,12,14,15]. In order to assess how weighting might influence the overall score, a detailed evaluation of the effect of weighting impact categories, or grouping impact categories into local and global impacts and weighting these, was undertaken. The results of this investigation demonstrated little significant effect on the relative scores for the benchmark data set and it was therefore decided to account for equal weighting amongst all categories by means of using an arithmetic mean. It should be emphasised that the methodology described above has been devised so that assessments undertaken by GSK scientists may be compared with existing GSK batch chemical processes used to synthesise APIs and the scores reflect this. However, it is a relatively simple matter to extend the upper score to reflect improvements in chemistry, technology and processes that enhance the life cycle profiles of 'typical' API chemical synthetic processes.

3.6 Complexity normalisation

The overall life cycle impacts of a pharmaceutical process are not only influenced by the complexity of the chemistries used in a synthetic route or process but also by the inherent molecular or chemical complexity of the API or intermediate being made. While this is not relevant when comparing different routes to the same drug, it is an important factor when attempting to compare and benchmark routes to different APIs. To account for differences in the molecular or chemical complexity of an API, one approach taken for this work involved using the molecular weight of the final product of the synthesis to normalise the scores for each category, as shown in Eq. 2. While this approach does not fully account for all the intricacies of molecular or chemical complexity, especially when encountering stereo-chemical and multi-functional molecular characteristics, it is a first step. Further enhancements to this approach are under active investigation and discussion.

3.7 FLASC™ scores: Interpretation

The FLASC™ score is a measure of the cradle-to-gate environmental life cycle impacts associated with the manufacture of materials used in the chemical synthesis of GSK's APIs or intermediates. A simple colour coding system is used to flag differences in scores.

- A 'green' rating is given for processes/routes with above-average performance (score ≥4). To achieve this, the life cycle impact associated with mass and energy will be <25% of the average for the benchmark data set.
- A 'red' rating is given for processes/routes with below average performance (Score ≤2). To achieve this, the life cycle impact associated with mass and energy will be >120% of the average for the benchmark data set.
- A 'yellow' rating is given for processes/routes with a score between 2 and 4.

Table 2 answers the question – 'what does a change in FLASC™ score mean in terms of the increase/decrease in overall environmental impacts?' For example, this means that for a typical process, an increase in the score from a '2' to a '3' equates to approximately a 50–60% reduction in the total environmental life cycle impact associated with the materials in a given process.

3.8 FLASC™

FLASC™ is a web-based tool and methodology available to GSK scientists and engineers. It delivers fast life cycle assessments of potential chemical synthetic routes or manufacturing processes used to make GSK APIs or intermediates. It also provides guidance about which materials have the greatest life cycle environmental impacts and allows a user to benchmark between existing or proposed routes or processes. Route or process assessment first requires that the name and quantity (kg/kg final product) of all materials used in the process be entered into FLASC™ via a simple spreadsheet. Process and material information are routinely generated by GSK R&D scientists from existing software systems or reports and FLASC™ has been developed to align with these existing formats.

Material Classification. If the LCI information for a material is contained within the database, the life cycle environmental impact data for that material will be automatically extracted. If the material is not in the database, a user must classify the material into one of the 14 material classes de-

Table 2: The FLASC™ score compared with the relative Life Cycle environmental impact for a given process

FLASC rating	% Relative to the average	Comments
5.0	12%	
4.3	20%	
4.0	25%	For a FLASC™ score = 4, the total life cycle mass and energy associated with the materials used is 25% of that associated with an average route.
3.8	30%	
3.4	40%	
3.1	50%	
2.9	60%	
2.7	70%	
2.5	80%	
2.4	90%	
2.3	**100%**	25 GSK routes developed during 1990 to 2000 were assessed. The average life cycle environmental impact was assigned a rating of 2.3.
2.1	110%	
2.0	120%	For a FLASC™ score = 2 the total life cycle mass and energy use associated with the materials is 120% relative to the average route
1.9	130%	
1.7	150%	
1.4	200%	
1.0	300%	For a score = 1 the life cycle mass and energy associated with the materials is 300% relative to the average route

scribed earlier. Material classifications are easily selected from a simple set of drop-down menus. In general, selection for a number of materials takes only a few minutes to complete. In those instances where classification is uncertain, a user may seek advice on classification by using the feedback facility on the GSK intranet site. The materials database will also be constantly reviewed to identify significant gaps, simplify classifications and updated as appropriate. Once all process materials are classified, FLASC™ will produce a final report. For each route or process, the report provides:

- the overall life cycle environmental impact score, and a breakdown for all impact categories;
- a summary of those materials having the largest life cycle net mass and energy use;
- data on reaction mass efficiency, mass productivity and solvent acceptability;
- appropriate guidance to help scientists make improvements.

Benchmarking and what-if scenario analysis is also possible. **Fig. 3** shows a typical comparison of several routes to a new product. Using the same approach it has been possible to assess FLASC™ scores for key 'GSK' APIs over a 20-year period, showing the significant benefits accruing from process improvement programmes.

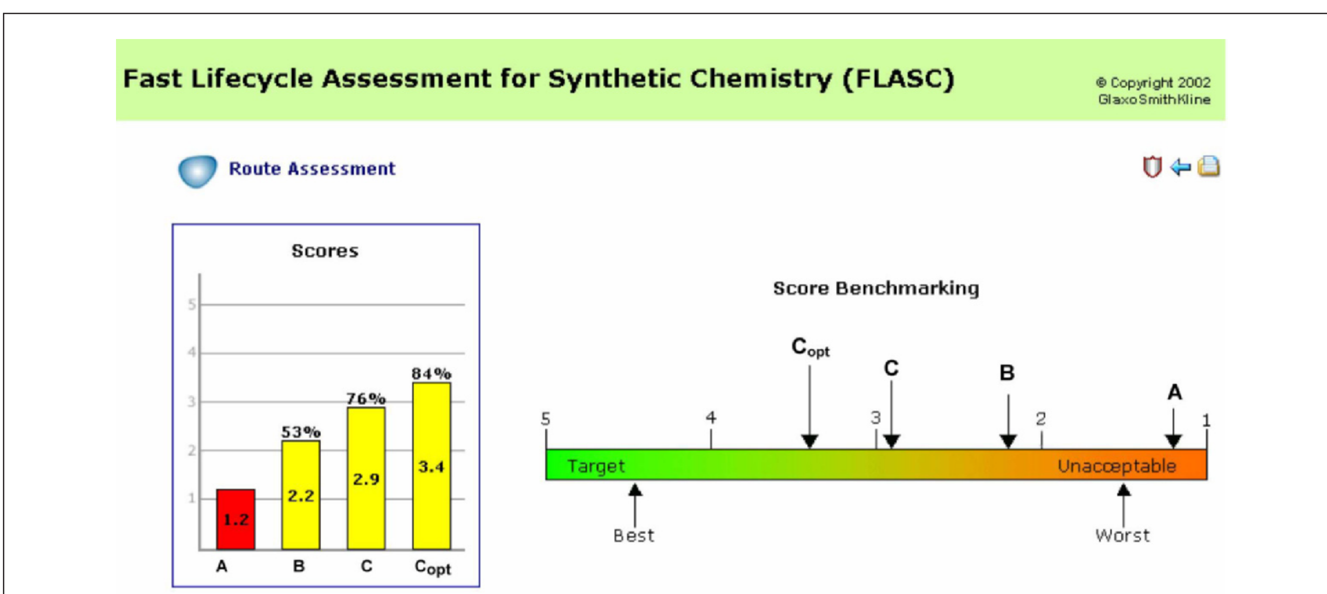

Fig. 3: Compares four different routes to an API, Routes A,B, C and C_{opt} (Route C optimized). FLASC scores are shown together with the percentage reduction of environmental impacts (compared to the worst route)

3.9 Validation and error analysis

Achieving an accurate and meaningful FLASC™ score requires the following:

- reliable, current and accurate LCI information;
- reliable, meaningful and accurate material classification;
- a valid and representative materials set in each of the 14 materials classes;
- an approach to take into account molecular complexity to enable meaningful comparison of routes to different products.

A number of variables were assessed during the validation process.

Error associated with use of average class data for new materials. For this methodology to be accurate, it is critical that there be a sufficient number and variety of materials with their associated life cycle data within each class. This will ensure a meaningful average impact data set that may be used for new materials. The number of materials in each of the 14 classes represented by this work varies from 3 to 20. Where there are few materials in a class, additional data will be developed.

Although it is true that cluster analysis will group materials with similar life cycle environmental impact data within a class, the variation in impact data within an impact category can be significant. Therefore the process of classification and use of average data for new materials, i.e., those that are not in the database of materials which possess actual life cycle data, will result in inaccuracies in the overall life cycle score.

To determine the significance of this potential error, a statistical analysis of the training data set was undertaken. First, the standard deviation (SD) of the individual environmental impact category data for the materials within each class was calculated for each of the 14 material classes. This SD was then used to calculate an overall SD for each of the 22 GSK processes that make up the benchmark data set. Typically, for any given process, there were between 2 and 8 materials that required the use of average data. The overall SD was generally found to be between 10–30%. In general, the higher standard deviations were associated with 'greener' processes; i.e., where the solvent usage was lower. However, the higher the score the smaller the effect of this error.

To further validate the model, an error analysis was performed by assessing the LCI profile for a GSK process for which all the LCI values were known [3] and these measured values were compared with output derived solely from the model (i.e., using each class' average). As a second test, the known values for all solvents were used and drew on the model only to supply the remaining compounds. The error between the 'observed' and the 'predicted' data was computed using the common statistical practice of taking the square root of the sum of the square weighted standard deviations. **Table 3** shows the results of the error analysis. It can be seen in Table 3 that when all the solvents are known, the expected error for most categories is less than 6%.

Because most of the solvent LCI data is in fact known, the small error is representative of the error that would be generally expected during common or typical use. In the case where all the materials are taken from the model, the error would be commensurately higher and would represent the worst-case scenario. Therefore, accuracy is maintained by two factors:

- for most GSK processes, solvents contribute >70% of the overall life cycle environmental impact [3]. Complete life cycle environmental impact data is now available within GSK for all GSK solvents in current use [13];
- the more complex a reactant or reagent (from a structural, functionality perspective), the greater the inaccuracy associated with the use of average life cycle data. To take account of this, FLASC™ users are asked to evaluate the synthesis of such materials back to relatively simple molecules and substitute these data. A potential source of the synthesis of these materials is the medicinal or discovery chemistry route that quite often is used to derive starting materials.

Sensitivity analysis has shown that for a typical GSK manufacturing process used to make an API, there are typically no more than 2 materials where classification is potentially ambiguous; e.g., complex organic or poly-substituted aromatic. Use of either usually made little difference to the score. However, this is not the case for very complex materials or for those materials based on natural products or fermentation and this is under evaluation.

4 Results and Discussion

4.1 An example of using FLASC™

To illustrate the application of this tool, a FLASC™ comparison of 4 different R&D synthetic routes/processes to the same API is presented in Figure 3. Route A corresponds to a close adaptation of the original (medicinal chemistry) route. Route B is a different route that incorporates some improvements to Route A. Route C is a significantly enhanced route derived from Route B, and Route C_{opt} resulted from optimising Route C during pilot runs.

Table 3: Results of the variation and error analysis. The table shows the percent in error measured by comparing results obtained with LCI known data and FLASC™ -generated results. The error was computed by taking the square root of the sum of the square weighted standard deviations

Data used for computation	Mass (kg)	Gross Energy (MJ)	POCP Equivs (kg et/kg)	GHG Equivs (kg CO_2/kg)	Acidification Equivs (kg SO_2/kg)	Eutrophication Equivs (kg PO_4^{3-}/kg)	TOC (kg)	Oil Equivs (kg)
Solvents known – Typical case	5.6%	5.9%	5.2%	6.1%	13.2%	26.1%	4.2%	4.6%
All unknowns – Worst case	40.8%	16.1%	32.0%	15.9%	25.7%	163.4%	40.1%	11.0%

The FLASC™ evaluation clearly demonstrates the improvement seen during the process development and identifies the process of choice. These data align with conventional EHS assessment data. In this particular example the optimal overall score for Route Cop is only 3.4 but the API is chiral and this necessarily adds extra complexity.

The output of FLASC™ illustrates not only the direct score and benchmark of the routes, but also shows the reduction in life cycle environmental impacts achieved through improvements in the synthesis and processes.

4.2 Value added to synthetic route improvement

The more efficient the route or process the better the resource utilisation (mass and energy) and the lower the associated impacts and cost. Two factors have the biggest impacts on the FLASC™ rating; namely, the:

- mass of materials used. This is influenced by the efficiency and complexity of the chemistry/technology and the manufacturing process;
- impacts associated with the individual materials.

FLASC™ provides information on those materials that have the largest contributions from both perspectives. It also provides values for the reaction mass efficiency, mass intensity, mass productivity, and solvent acceptability, and these provide further insight into opportunities for process improvement as well as benchmarking.

RoadMap to Better Processes

FLASC™ includes a list of key screening questions to help identify additional opportunities for process improvement. These questions are intended to help the chemist focus on the steps that can be taken during development to reduce the life cycle impacts of the synthetic route. The underlying objective of these questions is to provide not only the score of the route assessed, but also some guidance on aspects that could be looked at to improve its life cycle impact. Examples of the questions include:

1. Can a material with a better life cycle impact profile be substituted for a material with a poor impact profile?
2. Is there a different starting material that can be used in the synthesis to reduce the complexity of the synthesis?
3. Can the solvent be recovered and reused? In-house? Externally?
4. Can several steps in the synthesis be carried out in a single solvent?
5. Is solvent use optimised?
6. Are there any solvent-replacement operations that would generate solvent mixtures that could be difficult to separate/recover by distillation? If so, could these be avoided?
7. Can process intensification be used?
8. Are all intermediate isolations necessary?
9. Are 'catalysts' being used in stoichiometric amounts?
10. Can the catalyst be recovered or regenerated?

4.3 Comparison of life cycle impacts with GSK gate-to-gate operational impacts

In a thorough assessment of 'process greenness' it is important to understand the environmental (and health and safety) impacts across the entire life cycle associated with the manufacture of an API. There are two distinct parts to this:

- the cradle-to-gate impacts associated with the manufacture of the materials used in the GSK process to make an API, as determined by FLASC™;
- the GSK gate-to-gate impacts associated with the manufacture of the API from these materials.

A detailed comparison of mass and energy data has been undertaken for 17 well- developed GSK processes. Results shown in Fig. 4 indicate that there is a reasonably good correlation between the two parts of the life cycle when comparing:

- mass of raw materials extracted from the earth and the mass intensity used in a process;
- energy required to make materials and the GSK process energy required to manufacture a drug from those materials.

One point worth mentioning from **Fig. 4** is that the supply chain energy seems to be significantly larger when compared to the process energy. These data suggest that excluding process energy in the evaluations might only incorporate a small error. However, this observation does not hold when one compares process mass with the mass associated with raw material extraction and processing. Ongoing assessments

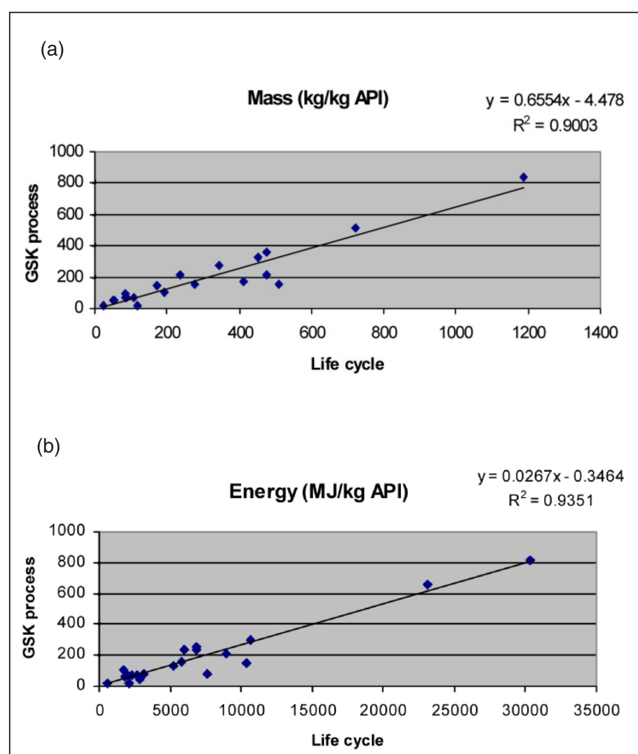

Fig. 4: Correlation between life cycle mass and GSK process mass [mass intensity] (a) and life cycle energy and GSK process energy (b)

with development processes exhibit the same type of correlation. It would therefore appear that from an environmental perspective, FLASC™ is an excellent indicator of process 'greenness' for cradle-to-post-GSK operations. Preliminary assessments also indicate that the FLASC™ score aligns with process economics.

It is not intended to directly assess process intensification, throughput, operability, scalability, waste or solvent recovery from GSK operations and currently does not incorporate specific chemical-related health or safety data. However, these aspects of process design are generally assessed within R&D and the results from FLASC™ are complementary to these evaluations.

5 Conclusions

The need for simple, but not simplistic, multi-functional environmental, health and safety (EHS) tools in an industrial setting is critical given decreases in EHS staff sizes and increased demands on workers' time and productivity. The work described here is an extension of GSK's philosophy for delivering innovative solutions to EHS issues through early intervention in the design of synthetic routes and chemical processes. The combination of tools now available to bench level scientists and engineers represents a significant resource for moving the company towards more sustainable business practices. Future work will continue to expand the utility of the toolkit and provide additional insights into materials and technology selection.

6 Recommendations and Perspectives

The following are being considered for future development of FLASC™.

- Generating and incorporating additional data to supplement categories where data is currently limited. A limited number of extra categories are being considered that will encompass processes based on fermentation and enzymation.
- Extending our evaluation of GSK processes to key new products and assessing GSK's development compounds using FLASC™ in a regular, milestone-driven basis.
- Using the methodology as a benchmarking tool for all parts of the corporation (i.e., not just for activities in GSK Pharmaceuticals R&D) is under consideration.
- Evaluating a framework for integrating health and safety data into life cycle assessments.
- Including evaluation of internal process energy into FLASC.
- Extending the evaluation and characterizations to natural and fermentation products.
- Comparing and correlating FLASC™ scores with other GSK metrics including Mass Productivity (MP), Reaction Mass Efficiency (RME) and Solvent Acceptability (SA).

Acknowledgements. The authors wish to acknowledge Brian Rohrback of Infometrix for his assistance with Pirouette and his collaboration during the statistical analysis. The authors also wish to acknowledge the significant contributions of GSK staff in Corporate Environment, Health and Safety; Corporate Information Technology (Jim McCann and Ben Holladay); and Pharmaceuticals R&D Chemical Development (John Hayler and Luisa Freitas dos Santos).

References

[1] Curzons A, Constable D, Mortimer D, Cunningham V (2001): So you think your process is green, how do you know? Using Principles of Sustainability to determine what is green – A corporate perspective. Green Chemistry 3, 1–6
[2] Constable D, Curzons A, Cunningham V (2002): Metrics to 'green' chemistry – Which are the best? Green Chemistry 4, 521–527
[3] Jiménez-González C, Curzons A, Constable D, Cunningham V (2004): Cradle-to-Gate Life Cycle Inventory and Assessment of Pharmaceutical Compounds: a Case-Study. Int J LCA 9 (2) 114–121
[4] Jiménez-González C, Kim S, Overcash M (2000): Methodology of developing gate-to-gate life cycle analysis information. Int J LCA 5 (3) 153–159
[5] Jiménez-González C (2000): Life cycle assessment in pharmaceutical applications. Ph.D. Thesis, North Carolina State University, 257pp
[6] US Department of Transportation, 1997 Economic Census. Transportation. 1997 Commodity Flow Survey (1999): US Department of Commerce, Economics and Statistics Administration. US Census Bureau. US Department of Transportation. Bureau of Transportation Statistics. EC97TFC-US
[7] Pirouette, Infometrix, Inc., P.O. Box 1528, Woodinville, WA <http//www.infometrix.com>
[8] Malinowski E (1991): Factor analysis in chemistry. Second Edition. John Wiley & Sons, Inc., New York
[9] Massart D, Vandeginste B, Deming S, Michotte Y, Kaufman L (1988): Chemometrics: A textbook. Elsevier, New York
[10] Breen J, Robinson P (eds) (1985): Environmental applications of chemometrics. American Chemical Society, Washington, DC
[11] Curran M (1996): Environmental life cycle assessment. McGraw Hill, New York, NY, USA
[12] SETAC (1993): A conceptual framework for life cycle impact assessment. SETAC foundation for environmental education inc., Pensacola, FL, USA, 105pp
[13] Jiménez-González C, Curzons A, Constable D, Cunningham V (2005): Expanding GSK's solvent selection guide – Application of life cycle assessment to enhance solvent selections. Clean Technology and Environmental Policy 7 (1) 42–50
[14] International Organization of Standardization (1997): ISO 14040 – Environmental management – Life cycle assessment – Principles and framework. ISO 14040:1997(E), Geneva, Switzerland, 12 pp
[15] International Organization of Standardization (2000): ISO 14042 – Environmental management – Life cycle assessment – Life cycle impact assessment. ISO 14042:2000(E), Geneva, Switzerland, 16 pp

Received: September 1st, 2005
Accepted: March 28, 2007
OnlineFirst: March 29th, 2007